

EJERCICIO PRÁCTICO 6: INFERENCIAS CON FRECUENCIAS – EL CASO DE LAS PROPORCIONES

CONTEXTO

Ya conocemos bien el contraste de hipótesis sobre la media de una población o las diferencia entre las medias de dos poblaciones. Incluso tenemos herramientas para buscar un nivel de significación (α) y un poder estadístico ($1 - \beta$) específicos.

Pero hay muchas preguntas que no se tratan sobre medias, puesto que involucran variables categóricas. Por ejemplo ¿cuánta gente está a favor de la ley de aborto?, ¿qué lista gana la próxima elección del CEII? ¿Cómo podemos hacer frente a este tipo de preguntas?

OBJETIVOS DE APRENDIZAJE

1. Entender diferentes casos de inferencia estadística con proporciones.
2. Realizar, utilizando el entorno R, inferencias o diseño de estudios con proporciones.

ÉXITO DE LA ACTIVIDAD

1. El equipo interpreta adecuadamente las preguntas planteadas, identificando el tipo de inferencia o diseño que se consulta.
2. El equipo construye un script en el entorno R para responder adecuadamente cada una de las preguntas planteadas.

ACTIVIDADES

Los siguientes datos se basan en un artículo publicado por Hart & Perlis (2019) (JAMA Internal Medicine, 179(9), 1285-1287) acerca de la proporción de mujeres autoras de artículos científicos en el área médica. La tabla muestra la cantidad de autoras y autores para diferentes especialidades.

Especialidad	Mujeres	Hombres
Pediatría	54	52
Obstetricia	71	66
Dermatología	35	41
Psiquiatría	30	42
Medicina Interna	45	65
Oncología	44	62
Neurología	56	88
Anestesiología	21	40
Radiología	17	35

1. Copiar los enunciados de los problemas asignados como comentarios de un script R.
2. Identificar qué se pide determinar en cada pregunta, justificando explícitamente con comentarios en el script.

3. Escribir código R para aplicar la solución propuesta, comentando el procedimiento adecuadamente.
4. Redactar respuestas a las preguntas planteadas (comentarios) en base a los resultados del análisis realizado.

Fuera del horario de clases, cada equipo debe subir el script realizado UVirtual con el nombre "EP06-respuesta-grupo-i", donde i es el número de grupo asignado. Las respuestas deben subirse antes de las 23:30 del sábado.

PREGUNTAS

Grupo 1:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de dermatología era de 61%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de oncología y medicina interna?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de obstetricia y la de neurología es de 0,19. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 97,5% y poder estadístico de 75%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 2:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de obstetricia era de 59%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de neurología y obstetricia?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de psiquiatría y la de neurología es de 0,25. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 95% y poder estadístico de 75%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 3:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de psiquiatría era de 48%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de pediatría y radiología?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de radiología y la de anestesiología es de 0,32. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 99% y poder estadístico de 80%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 4:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de pediatría era de 35%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de anestesiología y pediatría?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de medicina interna y la de dermatología es de 0,15. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 97,5% y poder estadístico de 80%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 5:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de medicina interna era de 42%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de obstetricia y radiología?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de anestesiología y la de pediatría es de 0,28. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 95% y poder estadístico de 80%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 6:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de oncología era de 32%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de oncología y dermatología?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de psiquiatría y la de obstetricia es de 0,18. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 99% y poder estadístico de 90%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 7:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de neurología era de 32%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de anestesiología y obstetricia?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de psiquiatría y la de medicina interna es de 0,23. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 97,5% y poder estadístico de 90%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 8:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de radiología era de 50%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de radiología y anestesiología?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de pediatría y la de dermatología es de 0,35. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 95% y poder estadístico de 90%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 9:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de anestesiología era de 57%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de medicina interna y pediatría?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de oncología y la de medicina interna es de 0,21. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 95% y poder estadístico de 85%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Grupo 10:

1. Estudios previos habían determinado que la proporción de autoras en la especialidad de pediatría era de 61%. ¿Respaldan estos datos tal estimación?
2. Según estos datos, ¿es igual la proporción de autoras en las áreas de obstetricia y oncología?
3. Suponiendo que la diferencia en la proporción de autoras en la especialidad de radiología y la de dermatología es de 0,29. ¿A cuántos autores deberíamos monitorear para obtener un intervalo de confianza del 99% y poder estadístico de 85%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

CRITERIOS DE EVALUACIÓN

Pregunta	Categoría	Nivel de logro	Puntos	Ideal
1	Hipótesis y prueba	Formulan con claridad y explícitamente hipótesis nula y alternativa adecuadas para responder la pregunta planteada sobre una proporción, tanto en lenguaje natural como en lenguaje matemático, identificando clara y explícitamente la prueba que puede contrastarlas	3	3
	Condiciones y aplicación	Escriben código R que verifica, o donde se argumenta convincentemente, el cumplimiento de las condiciones de la prueba, y que realiza de forma completa y correcta la prueba seleccionada con los datos correctos	3	3
	Conclusión	Entregan una conclusión correcta y completa a la pregunta planteada, basándose en el resultado de la prueba realizada y el contexto del problema	2	2
2	Hipótesis y prueba	Formulan con claridad y explícitamente hipótesis nula y alternativa adecuadas para responder la pregunta planteada sobre la diferencia entre dos proporciones, tanto en lenguaje natural como en lenguaje matemático, identificando clara y explícitamente la prueba que puede contrastarlas	3	3
	Condiciones y aplicación	Escriben código R que verifica, o donde se argumenta convincentemente, el cumplimiento de las condiciones de la prueba, y que realiza de forma completa y correcta la prueba seleccionada con los datos correctos	3	3
	Conclusión	Entregan una conclusión correcta y completa a la pregunta planteada, basándose en el resultado de la prueba realizada y el contexto del problema	2	2
3	Identificación	Reconocen que se pregunta por el tamaño de dos muestras, identificando claramente el contexto de la hipótesis nula que se desea contrastar, la prueba que necesita realizarse y los elementos conocidos (alfa, beta, tamaño del efecto, etc.)	4	4
	Tamaño de la muestra	Obtienen correctamente los tamaños de las muestras para una prueba de diferencia de proporciones, considerando los elementos conocidos (alfa, beta, tamaño del efecto, etc.) y manteniendo aproximadamente la proporción de personas estudiadas en cada caso	4	4
Todas	Datos	Obtienen los datos adecuados para responder cada una de las preguntas del ejercicio, identificando las frecuencias involucradas y determinando correctamente los totales y proporciones involucradas	3	3
	Calidad del código	Escriben código R -ordenado, bien indentado, sin sentencias espurias y bien comentado- que intenta dar respuesta a todas las preguntas planteadas	3	3
	Calidad de la escritura	Escriben con buena ortografía y redacción (≤ 3 errores), usando vocabulario propio de la disciplina y el contexto del problema	2	2
TOTAL				32
NOTA				7,0